Principled probability in syntax: A Minimalist approach to variability and cumulativity in phrasal movement

Overview The relevance of probabilistically variable patterns has long been controversial in linguistic theory, historically led by two dominant schools. On the first view, which I call PRINCIPLES WITHOUT PROBABILITY, variable patterns are outside the purview of formal theories of grammar, composed solely of categorical principles (Newmeyer 2003). The second view, PROBABILITY WITHOUT PRINCIPLES, holds that unit and collocation frequencies form the basis of mental representations, and denies the existence of formal categories (Bybee and Hopper 2001). I argue that these approaches should be reconciled in a theory with PRINCIPLES AND PROBABILITIES: Language is mentally represented as a generative system with formal categories and principles, as well as probability distributions over possible structures. I illustrate this approach with an analysis of word order variability in Cherokee that integrates Minimalist derivations with constraint-based optimization (Heck and Müller 2003), specifically the probabilistic computation of optimality of Maximum Entropy Harmonic Grammar (Goldwater & Johnson 2003). This approach expands the predictive and explanatory power of generative Minimalism to a broader range of word order phenomena, while maintaining its insights about the feature-driven nature of movement.

Probabilistic variability and cumulativity in Cherokee Cherokee allows a high degree of optionality in the placement of non-pronominal arguments (NPs) relative to verbs; all ordering permutations of agent NPs, theme NPs, and verbs are grammatical, as shown in the representative subset of examples in (1), ex. from Feeling et al. (2017: 101, 43, 129, 35).

(1) a. gitli ogi-sdawadvs-v
dog 1.PL.EXCL-follow-EXPP
'The dog followed us.' [***Ag.>V***]

b. a-n-adasdelis-g-o yvwi j-u-n-asdi
3-PL-help-PROG-HAB people DST-3-PL-little
'The little people help (others)' [***V>Ag.***]

c. am ji-todis-g-o
water 1-heat.water-PROG-HAB
'I heat water' [***Th.>V***]

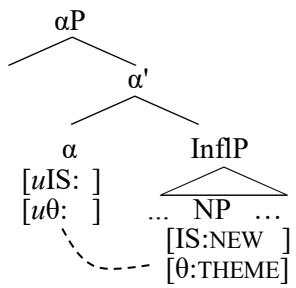
d. u-sdu-hnv galohisdi?i
3-close-EXPP door
'(he) closed the door.' [***V>Th.***]

The corpus study in Hsu & Frey (2024) shows that while some aspects of clausal word order are categorically predictable (frame-setting and constrast-bearing XPs are clause-initial), the placement of other non-pronominal arguments is *probabilistically* determined by two properties of NPs: its referential accessibility (i.e. givenness) and its thematic role. These factors contribute *cumulatively* to determine the order of argument NPs and verbs. The cumulative interaction of these preferences (discourse-new NPs are more likely to precede verbs than discourse-given NPs; agent NPs are more likely to precede verbs than theme NPs) is shown in the cross-pair table at the right for one set of features in clauses with one NP (the full paper discusses a broader range of accessibility and thematic-role values). NPs that refer to discourse-new agents are likelier to precede verbs than NPs with only one of either property. Crucially, NP placement is probabilistically variable, while holding constant the formal content of the numeration (grammatical roles, thematic roles, referential accessibility, etc.), contra the assumption that competing syntactic structures necessarily express distinct meanings (Newmeyer 2003: 697)

	Agent NP	*Theme NP*
New NP	**92%** preverbal (12/13)	**73%** preverbal (36/49)
Given NP	**76%** preverbal (37/49)	**50%** preverbal (57/114)

Clause structure and agreement I analyze preverbal vs. postverbal NP placement in terms of a movement operation that applies probabilistically, within the clause structure in Miyagawa (2009). Postverbal NPs occupy an (adjoined) position within InflP (Baker 1996); Non-contrasted preverbal NPs variably move to Spec, αP; Contrast-bearing NPs obgligatorily move higher to Spec, CP; verbs head-move to α. The interaction between information-structural and argument-related features in determining whether a non-contrasted NP moves to Spec, αP results from the co-location of two types of probes on α: I assume that NPs carry valued features corresponding to their assigned thematic role [θ:] (Hornstein 1999) and referential accessibility value [IS:].

(2)



These agree with corresponding probes on α, [*u*θ:] (independently able to account for effects of relative thematic prominence on agreement morphology on transitive verbs; Montgomery-Anderson 2015) and [*u*IS:]. Figure (2) shows agreement and valuation in a clause with one discourse-new, theme argument NP.

Violable constraints, probabilistic movement The probabilistic and cumulative aspects of movement triggering in Cherokee are problematic for the standard Minimalist view that the ability to trigger movement is an inherent property of individual probing features (i.e. *strength*; Chomsky 1993). In contrast, I show that gradient cumulativity in movement-triggering can be captured in a version of Minimalism where the outcome of each derivational step is determined by constraint interaction (Heck and Müller 2003), if optimality is calculated as a probability distribution over candidates, as in MaxEnt HG (Goldwater and Johnson 2003). The types of structures that this model can generate do not differ from standard Minimalism; candidate spaces are delimited by the set of syntactic operations, and the constraint inventory consists of grounded well-formedness restrictions on structure.

Phrasal movement satisfies FEATURE CONDITION (FC) constraints (Heck and Müller 2003); the grammar contains indexed versions of the constraint, corresponding to each value that a probe can receive, as in (3-4). Each constraint has a distinct, language-specific weight.

(3) FEATURE CONDITION (NEW): For each probe [*u*IS:] that agrees with an XP with [IS:NEW], the XP occurs in the specifier of the head that contains the probe.

(4) FEATURE CONDITION (AGENT): For each probe [*u*θ:] that agrees with an XP with [θ:AGENT], the XP occurs in the specifier of the head that contains the probe.

In the step after agreement, the grammar generates output candidates in which a phrasal movement occurs (satisfying one or more FC constraints), and candidates with no movement (violating one or more FC constraints). The tableaux in (5) show relevant candidates generated in the step after the [*u*IS:] and [*u*θ:] probes on α are valued by Agree. In MaxEnt HG, candidates with higher harmony (less penalty) are more likely to be chosen as optimal outputs, but not categorical winners. I show that a MaxEnt learner (Hayes and Wilson 2008) with minimal assumptions acquires a set of weights that generates the Cherokee pattern. The tableaux show the learned weights (*w*), harmony scores (*H*), and predicted probabilities (*P*) that correspond to the feature combinations in the first row of the cross-pair table on the first page. The potential for languages to vary in the weights of FC constraints predicts attested cross-linguistic variation in ordering preferences (ex. languages where discourse-given NPs preferentially precede discourse-new NPs).

(5)

$[_{\alpha'}$ α $[_{InflP} \ldots NP_{[IS:NEW][\theta:AGENT]}]$ [*u*IS:NEW] [*u*θ:AGENT]	FC(New) *w*=1.04	FC(Agent) *w*=1.16	*H*	***P***
$[_{\alpha P}$ NP $[_{\alpha'}$ α $[_{InflP} \ldots$ ~~NP~~] *movement occurs*			0	**.90**
$[_{\alpha P}$ α $[_{InflP} \ldots$ NP] *no movement*	-1	-1	-2.2	**.10**

$[_{\alpha'}$ α $[_{InflP} \ldots NP_{[IS:NEW][\theta:THEME]}]$ [*u*IS:NEW] [*u*θ:THEME]	FC(New) *w*=1.04	FC(Agent) *w*=1.16	*H*	***P***
$[_{\alpha P}$ NP $[_{\alpha'}$ α $[_{InflP} \ldots$ ~~NP~~] *movement occurs*			0	**.74**
$[_{\alpha P}$ α $[_{InflP} \ldots$ NP] *no movement*	-1		-1.04	**.26**

I then extend the analysis to word order tendencies in clauses with two NP arguments: Hsu & Frey (2024) find that the placement of each NP is determined by the same factors and probabilities observed for NPs that are the sole argument of a verb. I show that this is predicted if [*u*IS:] and [*u*θ:] agree with each NP argument in the clause; the placement of each argument NP is determined by the interaction of the same constraints and constraint weights. I argue that this Multiple Agree operation finds independent support in its ability to model portmanteau agreement on transitive verbs (Oxford 2019), which occurs in Cherokee.

References

Baker, Mark C. 1996. *The polysynthesis parameter*. Oxford: Oxford University Press.

Bybee, Joan L, and Paul J Hopper. 2001. Introduction to frequency and the emergence of linguistic structure. In *Frequency and the emergence of linguistic structure*, ed. Joan L. Bybee and Paul J. Hopper, 1–24. Amsterdam: John Benjamins.

Feeling, Durbin, William Pulte, and Gregory Pulte. 2017. *Cherokee narratives: a linguistic study*. Norman: University of Oklahoma Press.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, ed. Jennifer Spenader, Anders Erkisson, and Östen Dahl, 111–120. Stockholm: Department of Linguistics, Stockholm University.

Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39: 379–440.

Heck, Fabian, and Gereon Müller. 2003. Derivational optimization of wh-movement. *Linguistic Analysis* 33: 97–148.

Hsu, Brian, and Benjamin Frey. 2024. Word order in Cherokee: Information structure, thematic structure, and variability. *Language* 100: e124–e155. doi:10.1353/lan.2024.a937196.

Miyagawa, Shigeru. 2009. *Why Agree? Why Move? Unifying Agreement-Based and Discourse-Configurational Languages*. Cambridge, MA: MIT Press.

Montgomery-Anderson, Brad. 2015. *Cherokee Reference Grammar*. Ph.D dissertation. Norman: University of Oklahoma Press.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79: 682–707.